CGIAR Research Program on
Climate Change, Agriculture and Food Security (CCAFS)

Metadata
Video Transcript

October 2013

These materials were produced for and with funding from the Climate Change Agriculture and Food Security Research Program of the Consultative Group on International Agricultural Research (CGIAR).

Metadata

Researchers are often confused by metadata and what exactly a metadata document should contain. In this short video we try to explain what we understand by metadata and show how your standard documentation can fulfil your metadata requirements.

What is Metadata?

So what exactly is metadata – well it is information to enable a researcher to fully understand a dataset. This is often referred to as "Data about Data".

Over recent years there have been attempts to standardise metadata requirements – these include the "Data Documentation Initiative" and the "Dublin Core Metadata Initiative" – there are also several metadata editors available. Although useful in themselves these initiatives are quite technical and to a typical researcher can seem quite daunting and off-putting. We will try a different approach.

What should be included?

Imagine you were presented with a typical dataset like the one shown here. In order to understand this dataset there are several questions you would need answering. This might be:

- Where did this dataset come from? What study does it belong to? Can I have details of the study?
- Where was the data collected and when?
- Who carried out the study?
- What does each column represent?
- Are these coded data or "real" values – if coded, what do each of the codes mean?
- How were the data collected?
- Etc.

Your metadata should address these questions.

Study Level

Some of the questions are study level questions so we would have one level of metadata to describe the study – we will refer to this as the study catalogue. Think in terms of what, when, where, who – so for example

- What is the study about?
- When was the study carried out – give start and end dates if known
- Where was it done – describe the location
- Who carried out the study – who were the individuals or organisations involved

Data Level

Other questions such as what does each column represent, are data level questions – we will refer to this level as the data dictionary.

For each variable in your dataset you should indicate

- the data type,
- include a variable label to briefly describe it,
- if it is a coded variable then value labels will be needed so users can interpret the data correctly –
- have you used any missing value codes and what are they.

Example Catalogue

Here we can see part of the study catalogue for the CCAFS Household Baseline Survey. This includes a brief description of the study. Also included would be a reference to other documentation in this case the training manual.

Training Manual

The training manual for the household baseline survey includes a full description of the study, how sites were selected, the sampling methods used, the implementation plan, plus details of the sections of the questionnaire and why each section has been included.

Questionnaire

In the questionnaire used for the survey we have included variable names, this alleviates the need to have a separate document with all the variable labels. We have also included many of the codes within the questionnaire – a couple of examples visible in this extract are the relationship of the respondent to the household head, and the household type.

Codebook

Where there are too many codes to fit onto the questionnaire we have included a separate codebook. Part of the codebook is shown here – this extract shows some of the crop codes that were used.

Yes but…

Okay you say, but what's the minimum I need to do to provide metadata – do I need to produce a separate document?

Well no, not necessarily. For the household baseline we have the study catalogue giving a summary description and we have the training manual giving a full description. This satisfies the requirements at the study level.

At the data level we have the questionnaire which includes the variable names and we have the codebook. If we had included any derived variables we would need a document describing the calculation methods for each of these variables but as there are no such variables included in our archived data the questionnaire and codebook together fulfil the requirements for metadata at the data level.

Summary

So think about the two levels – study catalogue and data dictionary. Ensure that the documentation you supply gives enough information for the user to understand the study and the data.

Appendix I – CCAFS Data Management Support Pack

This document is part of the CCAFS Data Management Support Pack produced by the Statistical Services Centre, University of Reading, UK. The following materials are available in the pack:

0. Data Management Strategy
    a. CCAFS Data Management Strategy
1. Research Protocols
    a. Writing Research Protocols – a statistical perspective
    b. Preparation of Research Protocols – Good Practice Case Study
    c. What is a Research Protocol, and how to use one (Video & Transcript)
    d. Details of what a Research Protocol should contain (Video & Transcript)
2. Data Management Policies & Plans
    a. Creating a Data Management Plan
    b. Data Management Plan (Video & Transcript)
    c. Example Data Management Activity Plan
    d. Example Consent Form
3. Budgeting & Planning
    a. Budgeting & Planning for Data Management
    b. ToR Data Support Staff
    c. Budgeting & Planning (Video & Transcript)
4. Data Ownership
    a. Data Ownership and Authorship
    b. Template – Data Ownership Agreement
    c. CCAFS Data Ownership & Sharing Agreement
    d. Data Ownership & Authorship (Video & Transcript)
5. Data & Document Storage
    a. Creating and Using a DDS
    b. DDS Introduction – (Video & Transcript)
    c. DDS Organisation – (Video & Transcript)
    d. DDS Ownership – (Video & Transcript)
    e. Introduction to Dropbox – (Video & Transcript)
6. Archiving & Sharing
    a. Archiving & Sharing Data
    b. Data and Documents to Submit for Archiving – a checklist
    c. MetaData
    d. Archiving & Sharing (Video & Transcript)
    e. Metadata (Video & Transcript)
    f. CCAFS HBS Questionnaire
    g. CCAFS HHS Code Book
    h. CCAFS Training Manual for Field Supervisors

7. CCAFS Data Portals
    a. Portals for CCAFS Outputs
    b. AgTrials Summary
    c. CCAFS-Climate Summary
    d. DSpace Introduction
    e. Introduction to Dataverse (Video & Transcript)
    f. Creating a Dataverse (Video & Transcript)
    g. Dataverse Study Catalogue
    h. CCAFS Dataverse (Video & Transcript)
8. Data Quality & Organisation
    a. Data Quality Assurance
    b. Guidance for handling different types of Data
    c. Transition from Raw to Primary Data
    d. Data Quality Assurance (Video & Transcript)
    e. Guidance for handling different types of data (Video & Transcript)
    f. Transition from Raw to Primary Data (Video & Transcript)